BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e25647
doi: 10.3897/biss.2.25647

OPEN ACCESS

Conference Abstract

# Management of Molecular Data in DINA with SeqDB

Keith Glen Newton‡, Satpal Bilkhu‡, Nazir El-Kayssi‡, Christian Gendreau‡, James A Macklin‡

‡ Agriculture and Agri-Food Canada, Ottawa, Canada

Corresponding author: Keith Glen Newton (glen.newton@gmail.com)

Received: 09 Apr 2018 | Published: 18 May 2018

Citation: Newton K, Bilkhu S, El-Kayssi N, Gendreau C, Macklin J (2018) Management of Molecular Data in DINA with SeqDB. Biodiversity Information Science and Standards 2: e25647.
https://doi.org/10.3897/biss.2.25647

## Abstract

Agriculture and Agri-Food Canada (AAFC) is home to numerous specimen and environmental collections generating highly relational data sets that are analyzed using molecular methods (Sanger and NGS). The need to have a system to properly manage these data sets and to capture accurate, standardized metadata over entire laboratory workflows has been a long-term strategic vision of the Biodiversity group at AAFC. Without robust tracking, many difficulties arise when trying to publish or submit data to external repositories. To even know what work has been carried out on individual collection records over a researchers career becomes a demanding task if the information is retrievable at all. SeqDB was built to resolve these issues by centralizing, standardizing and improving the availability and data quality of source specimen collection data that is being studied using molecular methods. SeqDB also facilitates integration with tools and external repositories in order to take the burden off researchers and technicians having to create adequate systems to track and mobilize their data sets, allowing them to focus on research and collection management.

The development of SeqDB aligns with agile development methodologies and attempts to fulfill rapidly emerging needs from genetics and genomics research, which can evolve and fade quickly at times or be without clear requirements. The success of SeqDB as an application supporting DNA sequencing workflows has put it in the same space as other monolithic architectures before it. As the feature set to support the application continues to increase, the number of software developers vs operations and maintenance staff is difficult to rebalance in our organisation. In an effort to manage the scope for the project

© Newton K et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and ensure we are able to continue to deliver on our mandate, the sequence tracking workflows of the application will become part of the DINA ecosystem ("**DI**gital information system for **NA**tural history data", https://dina-project.net). Other functions of SeqDB such as collections management and taxonomy tree curation, will be replaced with the DINA modules implementing these functions.

In order to allow SeqDB to become a module of DINA, it has been decided to refactor the application to base it on a Service Oriented Architecture. By doing so, all molecular data of SeqDB will be exposed as JSON API Web Services (JavaScript object notation application programming interface) allowing other modules, user interfaces and the current SeqDB application to communicate in a standardised way. The new architecture will also bring an important technology upgrade for SeqDB where the front end will eventually become a project in itself.

## Presenting author

James A Macklin